This paper is published as part of a PCCP Themed Issue on:

Nucleic Acid Simulations

Guest Editors: Modesto Orozco and Charles Laughton

Editorial

Perspectives

Papers

A detailed binding free energy study of 2 : 1 ligand–DNA complex formation by experiment and simulation†

Witcha Treesuwan,‡[a] Kitiyaporn Wittayanarakul,‡[b] Nahoum G. Anthony,[b] Guillaume Huchet,[a] Hasan Alniss,[b] Supa Hannongbua,[a] Abedawn I. Khalaf,[c] Colin J. Suckling,[c] John A. Parkinson[c] and Simon P. Mackay*[b]

Received 1st June 2009, Accepted 12th August 2009
First published as an Advance Article on the web 22nd September 2009
DOI: 10.1039/b910574c

In 2004, we used NMR to solve the structure of the minor groove binder thiazotropsin A bound in a 2 : 1 complex to the DNA duplex, d(CGACTAGTCG)2. In this current work, we have combined theory and experiment to confirm the binding thermodynamics of this system. Molecular dynamics simulations that use polarizable or non-polarizable force fields with single and separate trajectory approaches have been used to explore complexation at the molecular level. We have shown that the binding process invokes large conformational changes in both the receptor and ligand, which is reflected by large adaptation energies. This is compensated for by the net binding free energy, which is enthalpy driven and entropically opposed. Such a conformational change upon binding directly impacts on how the process must be simulated in order to yield accurate results. Our MM-PBSA binding calculations from snapshots obtained from MD simulations of the polarizable force field using separate trajectories yield an absolute binding free energy (-15.4 kcal mol^{-1}) very close to that determined by isothermal titration calorimetry (-10.2 kcal mol^{-1}). Analysis of the major energy components reveals that favorable non-bonded van der Waals and electrostatic interactions contribute predominantly to the enthalpy term, whilst the unfavorable entropy appears to be driven by stabilization of the complex and the associated loss of conformational freedom. Our results have led to a deeper understanding of the nature of side-by-side minor groove ligand binding, which has significant implications for structure-based ligand development.

Introduction

Ligands interacting with DNA have the ability to regulate the gene machinery at the most fundamental level of expression.[1–3] Through their selective association with a particular DNA sequence, such ligands may prevent protein binding and play a role in treating diseases that result from aberrant gene expression.[4–7] The development of minor groove binders (MGBs) proceeded from the observation that two natural antibiotics, netropsin and distamycin, bind to A and T containing regions of the minor groove by a combination of hydrogen bonding with the bases on the groove floor facilitated by their natural curvature. Replacement of *N*-methyl pyrrole (Py) with *N*-methyl imidazole (Im) enabled the accommodation of the G–NH_2 by hydrogen bonding.[8–11] A significant breakthrough in the field was the observation that a number of MGBs could bind in the minor groove as a 2 : 1 complex, in a side-by-side fashion with the heterocyclic rings stacking against each other.[12] Since then, MGBs have been prepared that can discriminate not only GC from AT, but GC from CG and AT from TA base pairs.[13–15] Whilst hydrogen bonding to the groove floor endows specificity for particular sequences, it is not the only driving force for association; lipophilic forces, particularly interactions with the sugar moieties that comprise the groove walls are also highly relevant.[16,17] Furthermore, Haq *et al.*[18] showed that the electrostatic interactions between Hoechst 33258 and DNA are not major components of the driving force for binding; they simply replace the electrostatic interactions that exist between the water and ions with the uncomplexed ligand and DNA species.[18] This balance between enthalpic and entropic contributions to binding is the subject of extensive research, and appears to vary with MGB structure and the binding sequence of the DNA.[19]

When developing new DNA-binding ligands, it is therefore important to understand the factors that work both in favor of and counter to ligand binding. The molecular recognition event can be understood through detailed structural analyses using X-ray crystallography and NMR spectroscopy, but the intricacies of dynamic phenomena within ligand–DNA complexes need to be probed by other techniques such as molecular dynamics. The energetics associated with the

[a] *Chemistry Department and Center of Nanotechnology, Kasetsart University, Bangkok 10900, Thailand*
[b] *Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 27 Taylor Street, Glasgow, UK G4 0NR. E-mail: simon.mackay@strath.ac.uk*
[c] *WestCHEM Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, UK G1 1XL*
† Electronic supplementary information (ESI) available: DNA step parameters: comparison between simulated and experimental structures. See DOI: 10.1039/b910574c
‡ Contributed equally to this work.

binding event can be elucidated through thermodynamic analysis using techniques such as isothermal titration calorimetry (ITC) and circular dichroism. Our approach to ligand design is conducted in cognizance of these issues. We have prepared a large library of MGBs made up from new heterocyclic and head–tail groups that seek to recognize both the hydrogen bonding capacity of the groove floor to achieve specificity, and to exploit the lipophilic nature of the groove walls to enhance affinity.[20–23] We have found that increasing the size of the heterocyclic *N*- or *C*-alkyl groups, if placed towards the tail of the MGB, can offset side-by-side binding with the effect of extending the reading frame of the ligand from four to six base pairs. The first, well-characterized example of this effect was our detailed NMR study of the DNA duplex d(CGACTAGTCG)$_2$ with thiazotropsin A (formyl-Py-Py-^{iPr}Th-DMAP, where ^{iPr}Th represents thiazole containing an isopropyl moiety and DMAP the dimethylamino-propyl tail, Fig. 1). Furthermore, the sulfur and isopropyl of the thiazole not only improved lipophilicity, but also ensured that the heterocyclic nitrogen was inward facing, which enhanced sequence reading by introducing a hydrogen bond between the exocyclic amine of G7 and the ^{iPr}Th nitrogen.

To relate these structural examinations to binding energies, we have experimentally investigated by ITC the thermodynamics of thiazotropsin A interacting with the same DNA sequence. In parallel, we have used explicit solvent molecular dynamics (MD) simulations to relate the structural dynamics of ligand–DNA interactions explored by NMR with the experimental thermodynamic parameters to provide a unique insight into the processes that determine molecular association in this system.

Whilst MD simulations can provide predictions of ligand–DNA interactions through absolute binding free energy calculations, it is challenging to achieve accurate results.[24–26] Thermodynamic integration (TI), free energy perturbation (FEP), linear interaction energy (LIE) or combined molecular mechanics Poisson–Boltzmann and surface area (MM-PBSA) methods constitute the four different approaches for free energy MD simulations which have been applied to, for example, netropsin, distamycin, Hoechst 33258 and DAPI.[26–28] Although results have been rather encouraging using FEP methods,[28] the technique is not applicable to the general case. Accurate absolute binding affinities have been obtained, but only through extremely high computational cost.[29] Among the approximate methods, the MM-PBSA methodology has also been used to determine free energies in a number of nucleic acid systems,[24,25,30–32] although results tend to overestimate the actual experimental value.[30] The advantage of the MM-PBSA approach is the speed, reproducibility, reliability and efficiency of the calculations, compared with the resource requirements of FEP, LIE and TI. The MM-PBSA/GBSA approach was first introduced by Srinivasan[31] and represents a post-processing method of evaluating binding free energies and absolute free energies of molecules in solution by the analysis of sets of structures collected by MD or Monte Carlo methods. Representative structures from the trajectory are post-processed with solvent and counterions removed to calculate the free energy (G) according to eqn (1):

$$\Delta G_{\text{binding}}^{\text{solvated}} = \langle G_{\text{complex}} \rangle - (\langle G_{\text{receptor}} \rangle + \langle G_{\text{ligand}} \rangle) \qquad (1)$$

In previous studies aimed at analyzing ligand binding to DNA, eqn (1) has been applied to just the trajectory snapshots from the complex, which assumes that snapshots of the receptor and ligand taken from the complex trajectory are of comparable free energy to their separate trajectories. However, when the receptor and/or the ligand undergo significant conformational changes upon binding, this approach is less reliable and separate trajectories need to be considered to gain free energy values approaching experimental data, as demonstrated by Reyes and Kollman's study of RNA–protein associations.[33] Separate trajectories can only be considered if there are experimental structures available for the individual components, otherwise simulations extending beyond the nanosecond timescale may not yield accurate structures and ultimately produce imprecise energetic values for the separate systems. In this article, we have compared binding free energies derived using the single trajectory approach (where only one simulation of the receptor–ligand complex is performed and the co-ordinates of the receptor and ligand are subsequently extracted from the simulation of the complex for the purpose of free energy calculations) and the separate trajectory approach (where three distinct simulations are performed to obtain co-ordinates of the complex, receptor and ligand) to establish whether significant conformational changes impact upon our binding calculations.

According to the previous work we have alluded to,[24,34] calculating binding free energies using the MM-PBSA method from MD simulations has employed non-polarizable force fields that tend to overestimate the experimentally determined values. Hence, our study also includes a comparison between calculated binding free energies obtained from simulations using the polarizable and non-polarizable AMBER force fields, which differ primarily in their treatment of electrostatic interactions. The ff03 is a non-polarizable force field which is itself a modification of the ff94[35] and ff99[36] force fields. AMBER ff02 is a polarizable force field that explicitly includes induction effects by the use of dipole polarizabilities on all atoms, and has the effect of significantly increasing computer time compared with ff03. The polarizable ff02 has been applied previously to explore protein–ligand interactions with limited success.[37] Although DNA itself has been simulated using ff02 and ff99 to compare the reorganization energies[38] and their ability to maintain its structural integrity,[39] binding free energy calculations for ligand–DNA interactions on trajectories generated using these force fields have not previously been performed in a comparative manner. We herein report the first attempt using MD simulations to obtain

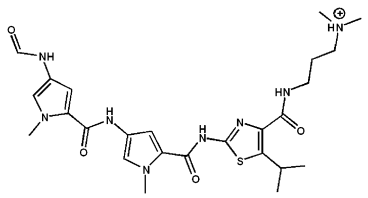


Fig. 1 Structure of thiazotropsin A.

absolute binding free energy data for 2 : 1 side-by-side binding of a ligand that evaluates the single and separate trajectory approaches using polarizable and non-polarizable force fields, all underpinned by structural and thermodynamic experimental assessment.

Methods

Isothermal titration calorimetry

Thiazotropsin A was prepared as before[20] and dissolved in degassed 0.01 M PIPES (piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)), 0.2 M NaCl, 0.001 M EDTA which had been adjusted to pH 7.0. d(GCGACTAGTCGC)$_2$ was purchased from MWG-BIOTECH AG (Anzinger str. 7a, D-85560 Ebersberg, Germany) as HPLC-purified salt free oligonucleotides synthesized on the 1 μmol scale. Ligand and DNA oligomer solutions were prepared in degassed buffer and the oligomer was heated to 90 °C for 12 minutes and allowed to anneal slowly over 12 hours. d(GCGACTAGTCGC)$_2$ was dissolved in 1 mL of the degassed PIPES buffer and the concentration of the resulting oligomer solution was determined spectroscopically at λ_{260} using the OD values supplied by the manufacturer. Aliquots were taken and diluted to achieve the concentration required for the ITC experiment ($\sim$15 μM).

ITC was performed at 25 °C using a Microcal VP-ITC (Microcal Inc., Northampton, USA). The control units were interfaced to PCs equipped with the Origin software package for data manipulation and instrumental control. The DNA concentration in the sample cell was 15 μM. Mixing was carried out by stirring the sample cell at 329 revolutions per minute. A 250 μL rotating syringe with an impeller profiled needle was used to perform 25 repeat 10 μL injections of the ligand with a 300 s delay between the first five injections, a 600 s delay between the subsequent fourteen injections and a 300 s delay between the last six injections. To correct for the dilution heat of the ligand, control experiments were performed at the same temperature using similar conditions with buffer only. The heats of ligand dilution were subtracted from the subsequent heat obtained for the titration of d(GCGACTAGTCGC)$_2$ with the ligand, thereby yielding the heat of binding for the ligand–DNA complex. All experiments were performed in duplicate.

MD simulations

The NMR structures of the free DNA decamer duplex d(CGACTAGTCG)$_2$ (Protein Data Bank code 1RN9) and in a 1 : 2 complex with thiazotropsin A (Protein Data Bank code 1RMX) were used for all simulations. A terminal GC base pair was added to either end of the DNA duplex to generate the dodecameric d(GCGACTAGTCGC)$_2$ that was used to determine the binding free energy experimentally by ITC. The AMBER 2003 (ff03)[40] and AMBER 2002 (ff02)[41] force fields that represent the non-polarizable and polarizable simulations, respectively, were applied to all DNA atoms. Since the RESP charges of thiazotropsin A are not available in the AMBER package, the RESP[42] methodology was applied to this ligand for our study. We initially optimized the thiazotropsin A monomer at the B3LYP/6-31G(d, p) level to adjust the structure obtained by NMR. The RESP fitting procedure was then used to obtain force field parameters for the ligand which was fitted by quantum mechanical electrostatic potentials at the HF/6-31G(d) level.

The separate trajectory approach used the co-ordinates for the free DNA solved experimentally (1RN9). The simulated systems were neutralized by the addition of 20 Na^+, 22 Na^+ and 2 Cl^- counterions for the complex trajectory (using the default AMBER parameters for these ions), free d(GCGACTAGTCGC)$_2$ and free ligand, respectively. Each system was placed in a periodic octahedral box solvated with TIP3P and POL3 water for the non-polarizable and polarizable force fields, respectively, with outer edges approximately 10 Å in each direction from the closest solute atom. Periodic boundary conditions with a 15 Å cutoff for non-bonded interactions were applied, with the particle mesh Ewald (PME) method[43,44] applied to account for the long-range electrostatic interactions.

Before the MD production phase, minimization and equilibration were carried out in three stages as follows: (i) the solvent and ions were minimized whilst the DNA and the ligand dimer were restrained by 10 kcal mol^{-1} Å^{-2} for 2000 steps, followed by heating using the NVT ensemble and Langevin dynamics with a collision frequency of 1 ps^{-1} from 100 to 300 K over 40 ps, followed by the NPT ensemble for 40 ps at a constant temperature of 300 K. (ii) Next, the restraints on the solvent and ions were applied at 10 kcal mol^{-1} Å^{-2} for the first 1000 steps and reduced to 5 kcal mol^{-1} Å^{-2} over 3000 steps of minimization using steepest descents and conjugate gradients, followed by equilibration from 100 to 300 K without pressure scaling. (iii) Minimizations were then performed again over a series of 1000 step intervals whilst restraints on the solute were gradually relaxed from 10, 5, 2, and 1 kcal mol^{-1} Å^{-2} and finally for 3000 steps without restraints. Further equilibration was applied to the completely unrestrained system using the NVT ensemble that involved heating from 100 to 300 K over 40 ps followed by NPT for 40 ps at a constant temperature of 300 K. The production phase involved the NPT ensemble at a constant temperature and pressure (300 K and 1 atm, respectively) for 5 ns using a timestep of 1 fs and the SHAKE[45] algorithm to constrain hydrogen to heavy atom bonds. The polarizable function was turned on throughout for systems that employed the polarizable force field. DNA structural parameters over the course of the trajectories were analyzed with CURVES.[46]

Binding free energy calculations

In principle, the MM-PBSA approach calculates free energies based on eqn (2). We used 100 snapshots of the solute sampled regularly from the last ns of the MD trajectories, with the water and counterions stripped away. This method combines the enthalpic or molecular mechanics energies (E_{MM}) that represent the internal energies (bond, angle and dihedral; E_{BADH}) along with van der Waals (E_{vdW}) and electrostatic interactions (E_{elec}), with the solvation free energies (G_{sol}) calculated by the finite difference Poisson–Boltzmann (PB) model for polar solvation (G_{PB} or G_{polar})[47] and the non-polar

contribution ($G_{\text{non-polar}}$) as a function of the solvent-accessible surface area (SASA). All terms were computed from the MM-PBSA module in AMBER. The conformational entropy (S), was approximated by normal mode analysis of harmonic frequencies calculated at the molecular mechanics level (eqn (2)).

$$G = \langle E_{\text{MM}} \rangle + \langle G_{\text{sol}} \rangle - T\langle S \rangle \quad (2)$$

The G_{polar} contribution was calculated by applying a cubic lattice with 0.5 Å grid spacing and evaluating all pairwise interactions using an internal dielectric constant of 1.0 and an outside dielectric of 80. The $\Delta G_{\text{non-polar}}$ was determined as a function of the SASA estimated using eqn (3), where γ and b are empirical constants of 0.0054 kcal mol^{-1} Å^{-2} and 0.92 kcal mol^{-1}, respectively, for water.

$$\Delta G_{\text{non-polar}} = \gamma \text{SASA} + b \quad (3)$$

Solute entropic contributions were estimated from the sampled structures based on normal mode analysis using the *nmode* module in AMBER. Due to demanding computational times, configurations were selected every 100 ps (water molecules and ions removed) from the last ns of the trajectory. The selected structures were minimized using conjugate gradients for 9900 steps after 100 steps of steepest descents. Newton–Raphson algorithms were then used for 5000 steps with a distance-dependent dielectric of $1/r^2$ (with r being the distance between two atoms) and a dielectric constant of 4 for the electrostatic interactions until the root-mean-square of the elements of gradient vector was less than 10^{-4} kcal mol^{-1} Å^{-1}. The frequencies of vibrational modes were computed at 300 K for these minimized structures using a harmonic approximation of the energies.

Binding free energies were determined by means of eqn (1) using snapshots from the last ns generated by the single and separate trajectory approach. For the former, co-ordinates for G_{receptor} and G_{ligand} were extracted from the G_{complex} trajectory. For the latter, snapshots were taken from the MD simulations performed separately on the solvated free DNA oligomer (using 1RN9 as the starting structure) and ligand dimer.

Results

Stability of the components in the single and separate trajectories

To assess the stability of the structures generated using the different protocols employed in our study for single (sg) and separate (sp) trajectories based on polarizable and non-polarizable force fields, RMSDs of the DNA oligomer–ligand complex (DNA_{CPX}), the DNA in single (DNAsg) and separate (DNAsp) trajectories were calculated with respect to the initial MD (NMR) structure and are shown in Fig. 2a and b, respectively. For both force fields, the complex as a whole and the DNA oligomer within the complex (DNAsg) displayed very stable trajectories, as indicated by the RMSD plots in Fig. 2. Greater, although acceptable, variations in RMSD were seen for the separate DNA oligomer trajectories (DNAsp) when the co-ordinates of the free DNA solved by

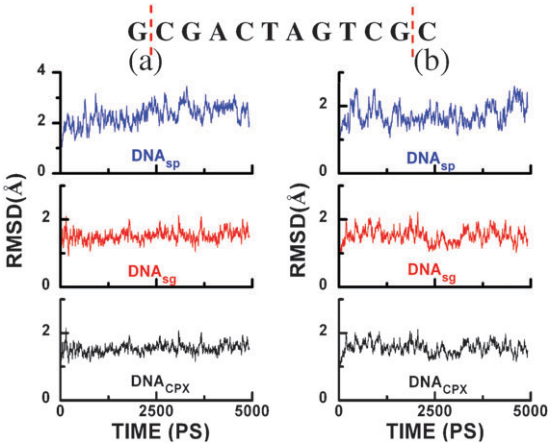


Fig. 2 All-atom root mean square deviation (RMSD) for simulations performed using the (a) polarizable and (b) non-polarizable AMBER force field. All-atom root mean square deviation of the complex (DNA_{CPX}-black); DNA from the single trajectory (DNAsg-red); and free DNA from the separate trajectory (DNAsp-blue). All RMSD calculations excluded the terminal nucleotides. Average RMSDs for the polarizable system were 1.53, 1.52 and 2.31 Å for the DNA_{CPX}, DNAsg, and DNAsp, respectively. Average RMSDs for the non-polarizable system were 1.55, 1.51 and 1.74 Å for the DNA_{CPX}, DNAsg, and DNAsp, respectively.

NMR were input into the simulations. Our results reflect those of Babin *et al.* who reported that dodecameric B-form DNA oscillated around an RMSD of 2.9 Å when a polarizable force field was used.[48] These RMSD values applied to all the system components (except for the dimer of thiazotropsin A when simulated in the free form (2LIGsp) for both force fields; Fig. 3) and indicate that stability had been achieved in the production phases. Furthermore, the heavy atom RMSD for each nucleotide (Fig. 4) of d(GCGACTAGTCGC)$_2$ when present in the complex (DNAsg), and when simulated separately (DNAsp), ranged from 0.45 to 0.58 Å for both systems, and demonstrated that the nucleotide parameters for both forms of DNA exhibited small differences for both force fields when compared with the NMR structures (Fig. 4).

The RMSD for the dimeric form of thiazotropsin A when simulated in the solvated unbound form (2LIGsp-nopol

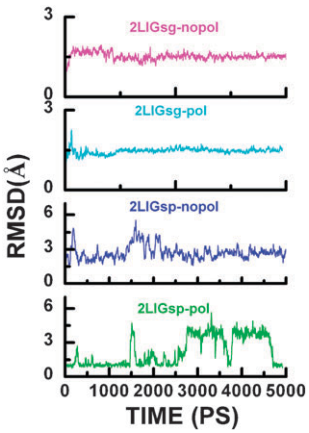


Fig. 3 RMSD for the dimeric form in the single (2LIGsg) and separate (2LIGsp) trajectories, when calculated for the polarizable (pol) and non-polarizable (nopol) force fields.

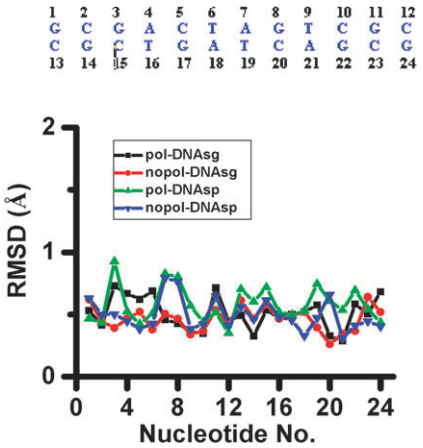


Fig. 4 RMSDs for the mean global heavy atoms of each nucleotide simulated using bound (DNAsg) or free (DNAsp) co-ordinates with the polarizable (pol) or non-polarizable (nopol) force field for 5 ns. Average RMSDs for the nucleotide heavy atoms in the polarizable system were 0.51 ± 0.1 and 0.58 ± 0.1 Å for the DNAsg, and DNAsp, respectively. Average RMSDs for the non-polarizable system were 0.46 ± 0.1 and 0.50 ± 0.1 Å for the DNAsg and DNAsp, respectively.

in Fig. 3) showed higher fluctuations using the non-polarizable force field than those for the polarizable force field for the first 2.5 ns. However, the opposite was observed for the last 2.5 ns. Furthermore, the free dimer favored two major conformations shown at an RMSD of about 4.0 and 1.0 Å from 2.5 ns to 5.0 ns for the polarizable force field. Whilst the solvated free form of the dimer does not directly impact on the structural integrity of our complex model, the presence of these two populations have implications for the free energy calculations (see Discussion).

Comparison of simulated structures with experimental data

To assess whether our MD simulations reproduced the experimental structural parameters we reported previously,[21] we conducted a detailed comparative analysis of the different force fields in the model systems.

DNA structural parameters. In order to compare the simulated and experimental structures in more detail than RMSD, key structural parameters of the free and bound DNA dodecamer were investigated. The step parameters that characterize the relationship between contiguous base pairs (*rise*, *roll*, *shift*, *slide*, *tilt*, *and twist*) over 100 snapshots from the last 1 ns production phase shown as the average values of all base pairs are reported in Table 1 (individual base pair parameters can be found in Fig. S1 and S2 in the ESI†). It appears that all step parameters for both force fields were consistent with the experimental data.

Minor and major grooves. The average groove width in the central -GACTAGTC- region for the bound DNA in all three systems were comparable (Table 2). A greater difference in groove width appeared between the simulated and experimental forms, which reflects the differences described by Laughton and Luisi,[49] who noted that groove features are affected by movement of the bases, which may enforce groove compression or expansion through a rolling action of the adjacent base pairs. These observations are also evident when comparing the step parameters of the free DNA forms (Table 1), particularly the *roll* parameter.

Ligand–DNA structural parameters

Hydrogen bonding. Hydrogen bond formation between the dodecamer and thiazotropsin A was determined based on a distance of 2.5 Å or less between the hydrogen bond donor (H) and acceptor (A) atom (H–A)—the empirical distance rule.[50] To investigate whether the simulated structures could reproduce the ten hydrogen bonds assigned to the experimental structure, the average distances in the polarizable and non-polarizable systems were measured (Table 3). In both systems, all hydrogen bonds detected by NMR were maintained, which demonstrated that our simulation protocols kept the integrity of the ligand–DNA complex within the parameters defined by experiment.

Inter-ligand distances. According to our experimental study, nuclear Overhauser effects (NOEs) indicated that there was little movement between the ligands not only relative to the DNA co-ordinates, but also with respect to one another. To demonstrate that the two ligands maintained this relationship within the simulated complexes, we showed that the distances between H4/C5 at the pyrrole methyl of LIG1 and C23/C24 of the isopropyl of LIG2 and *vice versa* for the polarizable and non-polarizable systems, respectively, remained consistent throughout the production phases (Fig. 5 and 6).

Free energy determinations using isothermal titration calorimetry

The thermodynamics of binding by thiazotropsin A to the dodecamer were examined using ITC to provide insight into the energetic basis for recognition and affinity by the ligand. The value of ΔH for a binding reaction is most reliably generated by calorimetry rather than indirectly from

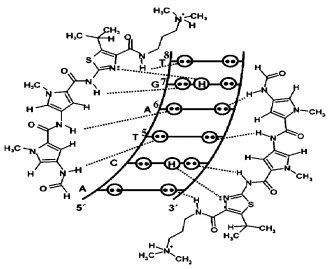
Table 1 The average step parameters for d(GCGACTAGTCGC)$_2$ in the complex with thiazotropsin A (CPX) and in the unbound form (DNA) calculated from the last ns of the trajectories for the polarized (pol) and non-polarized (nopol) force fields. The corresponding average step parameters for the experimental structures solved by NMR were computed using the average for all sets of co-ordinate supplied in the 1RMX and 1RN9 pdb entries

Parameter	CPX-pol	CPX-nopol	NMR-cpx	DNAsg-pol	DNAsp-nopol	NMR-DNA
Rise/Å	3.3 ± 0.3	3.3 ± 0.3	2.8 ± 0.2	3.4 ± 0.3	3.3 ± 0.3	3.1 ± 0.8
Roll/°	6.2 ± 4.8	6.2 ± 4.7	7.8 ± 4.1	3.9 ± 5.3	4.8 ± 5.1	1.4 ± 8.5
Slide/Å	0.0 ± 0.4	0.0 ± 0.5	−0.1 ± 0.3	−0.2 ± 0.6	−0.6 ± 0.6	−0.1 ± 0.4
Shift/Å	0.0 ± 0.5	−0.1 ± 0.5	0.5 ± 0.3	0.1 ± 0.6	0.0 ± 0.7	−0.9 ± 0.5
Tilt/°	0.2 ± 4.9	0.1 ± 4.4	0.5 ± 2.3	0.3 ± 4.3	0.3 ± 4.3	−0.4 ± 5.6
Twist/°	33.3 ± 3.9	33.9 ± 3.9	31.4 ± 3.5	30.5 ± 5.0	31.9 ± 5.0	33.5 ± 6.2

Table 2 Comparison of the minor and major groove width for bound and free DNA with the experimental NMR structure in the polarized (pol) and non-polarized (nopol) systems

	Minor groove width/Å			Major groove width/Å		
	NMR	Pol	Nopol	NMR	Pol	Nopol
Bound	7.0	7.0	7.7	11.7	13.6	11.9
Free	4.0	6.3	6.7	14.5	13.5	12.8

Table 3 The average distance between the hydrogen bond donor and acceptor atoms identified by experiment between ligand and DNA, and reproduced for simulated structures based on the polarizable and non-polarizable force field. Deviations throughout the simulations are shown in parenthesis



		Distances/Å	
Ligand atom	DNA atom	$DNA_{CPX\text{-}pol}$	$DNA_{CPX\text{-}nopol}$
H2	T^5O2	1.9(0.1)	1.9(0.1)
H9	A^6N3	2.7(0.6)	2.5(0.3)
H16	G^7N3	2.5(0.3)	2.4(0.2)
Thiazole N21	G^7H22	2.2(0.2)	2.1(0.2)
H26	T^8O2	2.6(0.4)	2.9(0.5)

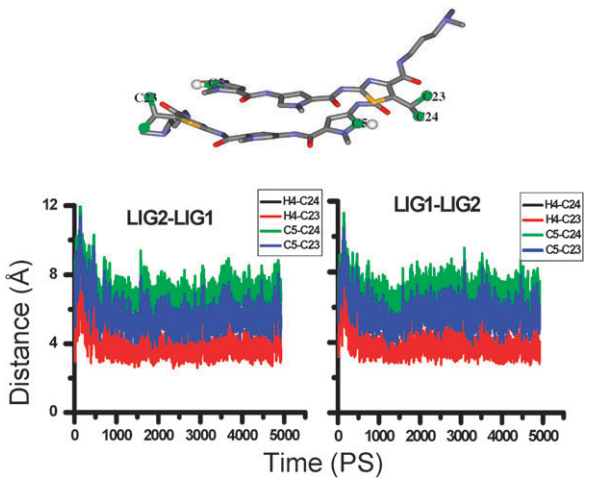


Fig. 5 Top: side-by-side anti-parallel conformation adopted by thiazotropsin A dimers with atoms involved in distance measurements annotated. Bottom: distances between atoms of the head group (H4, C5) and the isopropyl group (C23, C24) of the dimeric ligands (LIG1 and LIG2) in the bound complex (polarizable system).

van't Hoff determinations.[51,52] Titration of thiazotropsin A with the dodecamer in PIPES buffer at 25 °C was clearly associated with an exothermic process (Fig. 7A). Dilution peaks (data not shown) were all endothermic, and their intensity decreased as more ligand was added, indicating that aggregation of the

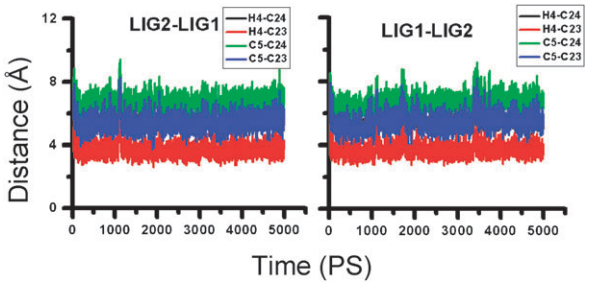


Fig. 6 The distances between atoms of the head group (H4, C5) and the isopropyl group (C23, C24) of the dimeric ligands (LIG1 and LIG2) in the bound complex (non-polarizable system).

thiazotropsin A occurs in buffered solution. The enthalpogram (Fig. 7B) was generated by integrating the raw data of the binding experiment and subtracting the heats of ligand dilution to yield binding enthalpy, ΔH, binding free energy, ΔG, and entropy changes, ΔS. Our analysis generated values for ΔG of -10.2 kcal mol^{-1}, ΔH of -12.9 kcal mol^{-1}, ΔS of -9.1 cal mol^{-1} K^{-1} ($T\Delta S = -2.7$ kcal mol^{-1}) and K of 3.0×10^7 M^{-1} and a binding stoichiometry of 2 : 1 (ligand to DNA). These results indicated that the binding interaction was enthalpically driven, and that there was an entropic penalty associated with the complexation process. Thiazotropsin A binding was measured at ligand : DNA ratios (r) that varied between 0 and 5.5 (Fig. 7), and analysis of the heat effects resulting from the binding process revealed that when $r \leq 2$, the binding enthalpy remained constant. A dodecamer can potentially provide two separate binding sites for a small molecule with dimensions similar to Hoechst 33258,[25] which raises the question: 'Does thiazotropsin A bind to two individual binding sites as a monomer or to one site as a dimer?' If monomeric binding occurred with relation to different base pair sequences of the two potential binding sites, a noticeable difference in the enthalpy of binding would be observed as the first and subsequently the second site was occupied. Since the measured ΔH values remained constant when $r \leq 2$, we propose that in this range, the ligand binds to the dodecameric DNA sequence exclusively in a non-stepwise dimeric 2 : 1 mode, which is consistent with our NMR study.

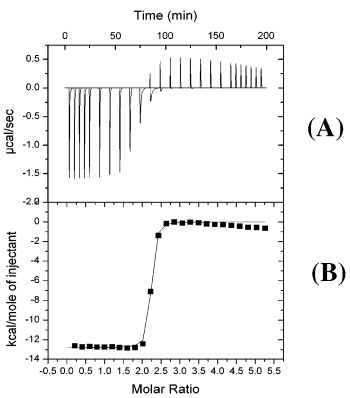


Fig. 7 (A) Raw data for titration of thiazotropsin A into the dodecamer in PIPES buffer at 25 °C (pH 7). (B) Enthalpogram retrieved from A, corrected for the heat of dilution, representing least-squares-fit to the single-site binding model.

Table 4 Adaptation energies of the d(GCGACTAGTCGC)$_2$ oligomer and the ligand dimer upon complexation using the polarizable and non-polarizable force fields. All values are in kcal mol^{-1}. Numbers in parentheses indicate the standard error. All terms are described in Methods

DNA/ligand	*E*(BADH)	*E*(vdW)	*E*(elec)	*E*(MM total)	*G*(PB)	*E*(elec) + *G*(PB)	G(non-polar)	G(total) + TDS
DNApol-bound	1111.7(18.9)	−188.5(10.1)	160.0(40.1)	1083.2(42.0)	−6244.8(37.3)	−6084.8(10.4)	27.0(0.2)	−4560.6
DNApol-free	1086.6(19.7)	−196.9(9.6)	153.3(56.3)	1043.0(56.5)	−6240.9(52.4)	−6087.6(10.6)	26.8(0.2)	−4602.6
Adaptation energy	25.1	8.4	6.7	40.2	−3.9	2.8	0.2	42.0
DNAnopol-bound	1116.0(22.6)	202.2(9.8)	443.7(32.7)	1357.5(33.9)	−6299.7(29.1)	−5856.0(12.8)	27.3(0.2)	−4359.1
DNAnopol-free	1102.0(19.2)	−206.1(10.6)	397.6(45.5)	1293.6(45.4)	−6255.6(41.8)	−5857.9(12.3)	27.2(0.2)	−4378.2
Adaptation energy	14.0	3.8	46.1	63.9	−44.2	1.9	0.1	19.1
LIGpol-bound	173.7(8.4)	−1.9(4.6)	−501.6(4.0)	−329.9(8.5)	−177.4(2.3)	−679.0(3.1)	8.3(0.1)	−355.0
LIGpol-free	182.0(10.1)	−9.8(3.8)	−514.5(10.3)	−342.3(12.5)	−175.6(8.8)	−690.1(3.7)	8.2(0.2)	−375.8
Adaptation energy	−8.4	7.9	12.8	12.4	−1.8	11.0	0.2	20.8
LIGnopol-bound	172.8(8.5)	−2.6(4.3)	−501.6(4.2)	−331.8(8.4)	−176.9(2.4)	−678.9(3.1)	8.4(0.1)	−366.4
LIGnopol-free	180.3(9.0)	−10.4(4.3)	499.5(11.5)	−329.6(14.2)	−187.2(10.7)	−686.7(3.4)	8.2(0.2)	−373.7
Adaptation energy	−7.5	7.9	−2.5	−2.3	10.3	7.8	0.2	7.3

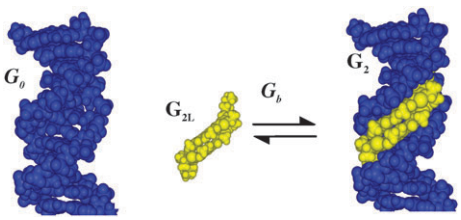
Free energy calculations using MD simulations

Adaptation energies of the bound and free forms of the complex components. Calculation of the adaptation free energy provides an indication of the conformational change and the energetic penalties involved when a system moves between a bound and free state. By this process, we can investigate whether a single or separate trajectory approach to determining the binding free energy is more appropriate. The single trajectory approach relies on minimal conformational rearrangement taking place when the free components associate into their bound complex and is the approach that has been predominantly used to evaluate ligands binding to DNA in the minor groove. When significant conformational adaptation within the DNA takes place, for example when ligands bind through intercalation or through an induced fit mechanism, the single trajectory approach fails to reproduce experimental binding free energies.[24,53,54] The adaptation free energy associated with binding was calculated using 100 snapshots from the last ns of each trajectory (for components in the free and bound states) according to eqn (2), followed by subtracting the energy of the free state from the bound state (Table 4).

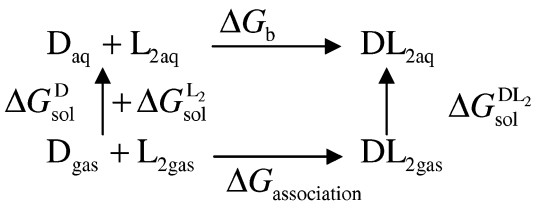
For the d(GCGACTAGTCGC)$_2$ dodecamer, the total average adaptation energies for the polarizable and non-polarizable systems were 42.0 and 19.1 kcal mol^{-1}, respectively, which implies that there is a significant penalty associated with conformational rearrangement from the free to the bound form using both force fields. Similarly, the ligand dimer undergoes a free energy penalty through conformational rearrangement on moving from the free dimer in solution to the bound dimer in the complex; the total average adaptation energies for the polarizable and non-polarizable force fields were 20.8 and 7.3 kcal mol^{-1}, respectively. Consequently, for ligand association with the DNA dodecamer to take place, the free energy of binding must offset these substantial conformational penalties.

Binding free energies. Our NMR studies indicate the presence of only the 2 : 1 complex structure during titration with the ligand; indeed, all our experimental studies (circular dichroism; capillary electrophoresis;[55] data not shown) have thus far found no evidence for the formation of an initial 1 : 1 complex prior to 2 : 1 binding. Furthermore, we have shown that thiazotropsin type ligands behave like aggregates of head-to-tail dimers in solution, and monomers are not present at the ligand concentrations under investigation (data not shown, manuscript in preparation). We therefore suggest that the reaction scheme follows the equilibration process described in Scheme 1, and the thermodynamic cycle in Scheme 2 represents the parameters that can be calculated using the MM-PBSA treatment of the respective solvated trajectories (eqn 4–6), irrespective of whether the determination uses the single and separate trajectory approach.

The ΔG_b of the single trajectories were −52.1 and −76.6 kcal mol^{-1} for the polarizable and non-polarizable force fields, respectively (Table 5). These clearly overestimate the experimental binding free energy of -10.2 ± 0.2 kcal mol^{-1}. Furthermore, for the separate trajectory approach using the non-polarizable force field, a ΔG_b value of −46.4 kcal mol^{-1} is still an overestimation of the real value. For the separate trajectory approach (Table 5), the appearance of two populations for the ligand dimer when simulated alone with the polarizable force field (see Fig. 3) complicates the calculation and introduces two conformational groups that can inform the ΔG_b of this system. We can describe two ΔG_b energies accordingly: (i) −16.2 kcal mol^{-1} over 4–4.6 ns. (ii) −14.5 kcal mol^{-1} over 4.8–5 ns, both of which approach the experimental result of −10.2 kcal mol^{-1} (the 4.6–4.8 ns interval is neglected because the conformations change dramatically over this period). The average ΔG_b (−15.4 kcal mol^{-1}) of these two conformations provides the closest approximation with the experimentally determined value.



Scheme 1 Schematic representation of the binding free energy of the duplex DNA d(GCGACTAGTCGC)$_2$ (blue) bound to the ligand dimer (yellow).



$$\Delta G_b = \Delta G_{association} + \Delta\Delta G_{sol} \quad (4)$$

$$\Delta G_{association} = \Delta H_{association} - T\Delta S_{association};$$

$$\Delta H_{association} = \Delta E_{MM}^{DL_2};$$

$$-T\Delta S_{association} = -T\Delta S^{DL_2} \quad (5)$$

$$\Delta\Delta G_{sol} = \Delta G_{sol}^{DL_2} - \Delta G_{sol}^{D} - \Delta G_{sol}^{L_2} \quad (6)$$

Scheme 2 Thermodynamic cycle for the association of the thiazotropsin A dimer (L_2) with d(GCGACTAGTCGC)$_2$ (D).

Discussion

Structural comparisons of simulated with experimental structures

A key element of any DNA simulation is the treatment of the long-range electrostatic interactions to ensure the stabilization of this highly polar and charged molecule. For biomolecules such as nucleic acids, a recognized source of inaccuracy in current force fields is the treatment of the electrostatic interactions, which has led to the development of polarizable force fields that are potentially more accurate than those based on point charges. However, these force fields have been largely unexplored with DNA, although extensive and detailed studies have started to emerge that examine their capabilities and limitations. In one such investigation, Babin *et al.*[48] compared polarizable and non-polarizable force fields based on the original Cornell *et al.*[35] description and found that the former tended to perform better at reproducing the structural features of DNA. Moreover, they were able to improve upon previous solution simulations,[56–58] reducing RMSDs from around 4 Å to below 2 Å and represent the best results achieved to date that can be obtained with present force fields in solution phase studies. In this study, it appears that the DNA co-ordinates in the complex showed lower deviation than the free DNA, which suggested a stabilizing influence by the bound dimer for both force fields. This was reflected in the RMSD of the dimer itself: when associated with DNA in the complex it exhibits minor deviations (Fig. 3; 2LIGsg-pol; 2LIGsg-nopol), which become more exaggerated when simulated in the free form (2LIGsp-pol; 2LIGsp-nopol).

To compare the simulated and experimental structures in more detail than RMSD, we examined several structural parameters of the free and bound DNA. Overall, both force fields were able to recreate stable trajectories that were structurally consistent with those we had solved previously by NMR and both performed equally for the simulations of the DNA–ligand complex. The RMSDs seen for the separate trajectories of the free DNA were more pronounced, but were still within the range reported in previous studies.[58] Unlike Babin *et al.*, greater oscillation was found for the free dodecamer in the polarizable system compared with the non-polarizable force field, although the average RMSD of 1.74 Å in the polarizable model generated here approximates to their value of 1.62 Å. The lower fluctuations that were observed throughout the bound complex trajectories replicate the known stiffening effect that minor groove binding ligands have on binding to nucleic acids,[26,59] whilst the larger fluctuations for the free DNA reproduce the inherent flexibility observed in experimental systems. Having established that both force fields were able to describe this system reliably from a structural perspective, investigations were then sought to establish whether these simulations could be translated with any accuracy into the experimentally determined binding thermodynamics that were obtained from the ITC measurements.

Binding free energies: comparisons of simulated with experimental structures

The titration of thiazotropsin A with the DNA dodecamer was clearly associated with an exothermic process (Fig. 7A) together with an endothermic dilution process indicative of aggregation of the thiazotropsin A prior to DNA binding. A favorable enthalpy of interaction (ΔH of -12.9 kcal mol^{-1}) is in agreement with the observation that exothermic interactions occur for the vast majority of ligands binding to DNA at room temperature.[19] Analysis of our binding isotherm resulting from the titration of thiazotropsin A revealed that when $r \leq 2$, the ΔH values remained constant for the binding of thiazotropsin A with d(GCGACTAGTCGC)$_2$. It can be concluded that the ligand binds exclusively in a dimeric 2 : 1 mode, which is consistent with our previous NMR study, and agrees with the detailed comparative study of the

Table 5 The molecular mechanics ($\Delta E_{MM}^{DL_2}$), solvation (ΔG_{sol}), entropic ($-T\Delta S^{DL_2}$), and binding free energy (ΔG_b) terms from eqn (4) to (6) for thiazotropsin A binding to DNA using the polarizable and non-polarizable force fields. All values are in kcal mol^{-1}. Conf. 1 and Conf. 2 represent the two major conformational populations occupied by the free dimer (2LIGsp) using the polarizable force field (Fig. 3). All_conf. is the mean value for the two populations

	Single traj.			Separate		
Energy component	Pol	Nopol	All_conf.	Pol Conf. 1	Conf. 2	Nopol
$\Delta E_{MM}^{DL_2}$	−1281.1	−1287.2	−1229.8	−1226.3	−1233.2	−1225.0
ΔG_{sol}	1183.8	1179.6	1180.2	1178.2	1182.2	1146.0
$-T\Delta S^{DL_2}$	45.2	31.0	34.2	31.9	36.5	32.6
ΔG_b	−52.1	−76.6	−15.4	−16.2	−14.5	−46.4
ΔG^{exp}			−10.2			

thermodynamics of distamycin and netropsin binding performed by Lah and Vesnaver.[60] Moreover, Lah and Vesnaver's study[60] also revealed that distamycin binding in a 2 : 1 fashion is characterized by a strong enthalpy of similar magnitude to thiazotropsin A (*e.g.* $\Delta H = -12.5$ kcal mol^{-1}) and is also accompanied by a substantial unfavorable entropy contribution (*e.g.* $T\Delta S = -2.0$ kcal mol^{-1}). They also showed that binding free energy was dominated by a combination of non-covalent interactions such as hydrogen bond formation and van der Waals interactions, and through the hydrophobic transfer of the ligand from the surrounding solution to its binding site within the DNA minor groove.

Calculating absolute binding free energies poses a major challenge, particularly when considering practical problems such as inadequate sampling and the need to make approximations. This is particularly evident with the MM-PBSA approach, and the rigid-binding approximation that involves the single trajectory protocol has traditionally been used to overcome the incomplete sampling limitations associated with the separate trajectory approach. Rapidly improving computational architectures have helped reduce sampling limitations, and based on the relative success of the MM-PBSA method with rigid-binding approximations, researchers have been applying the methodology to tackle systems of higher flexibility where binding results in small conformational changes, with varying degrees of success.[34,61–64] Reyes and Kollman,[33] when simulating an RNA–protein association, found that separate trajectories of the monomers, using their unbound experimental structures, provided the best agreement with the experimental binding free energies. Perhaps significantly, in the study by Spackova *et al.*[24] investigating the binding of the ligand DAPI to dodecameric DNA by the single and separate trajectory approach, the co-ordinates for the free DNA were taken from the simulated complex after removing the associated DAPI. These co-ordinates are appreciably different from those found in uncomplexed experimental free DNA, which could suggest why their flexible-binding approach was less successful than the rigid approach.

In order to establish whether significant conformational rearrangement took place when thiazotropsin A bound to the dodecamer, the adaptation energies of the nucleic acid and the ligand dimer upon binding were calculated as the difference between the free energy of the molecules from the simulation of the free state and the bound state.[65] Adaptation energy states that the receptor must undergo work to reorganize in order to accommodate the ligand, and the more positive the value, the greater the deformation required. By this process, it is possible to investigate whether a single or separate trajectory approach to determining the binding free energy is more appropriate (Table 4). For the DNA dodecamer, the adaptation energies of 42.0 and 19.1 kcal mol^{-1}, respectively, for the polarizable and non-polarizable force fields imply a significant penalty associated with binding. Overall, these values reflect the strain induced within the DNA structure by moving from the free to the bound state, which is only compensated in part by the solvation term (G_{PB}) that offers the only favorable drive. The cancelling of E_{Elec} with G_{PB} in both force fields appears to nullify the effects of the charged contributions to the overall energetics, leaving the greatest difference (E_{BADH}) arising from the parameter set that is common to both force fields. The subtle differences in structure that arise through the influence of the non-bonded electrostatic term must therefore account for the differences between both force fields.

In summary, for ligand association with the DNA to take place, the free energy of binding must offset the substantial conformational penalties described by the adaptation energies and suggests that a flexible-binding approach using separate trajectories should be employed to simulate the association.

Both force fields significantly overestimate the binding free energy using the single trajectory protocol, which we relate to the neglect of conformational changes upon binding.[24] The positive adaptation free energies for all components in the system upon complexation (Table 4) indicate the magnitude of the conformational change that takes place, and the offset required by the contributions to binding free energy to overcome this effect. We found that the binding free energy obtained from the separate trajectory approach approximated and incorporated these conformational changes to a better degree, and returned values more consistent with our experimental studies, particularly using the polarizable force field.

Having established that structures generated by the separate trajectory approach described by the polarizable force field and post-processed using the MM-PBSA methodology furnished binding free energy values closest to experiment, evaluation of the contributions that promoted and countered ligand binding were sought. ITC had revealed that the binding process of thiazotropsin A to d(GCGACTAGTCGC)$_2$ was principally enthalpy-driven, although countered by an unfavorable entropy contribution. From our simulation studies, we parsed the contributions to the free binding energy described by eqn (4) into the individual components described by eqn (2), which were then broken down further into the bonded (BADH) and non-bonded (VDW; ELEC) terms and the polar (PB) and non-polar (SA) contributions. This was in order to establish where the driving force for binding at the molecular level arises from (Table 6). Like our experimental determination, the entropic term was unfavorable and opposed the binding process. In our simulation, the entropy contribution arises from a normal mode analysis, and reproduces the entropic penalty on complex formation through the loss in translational, rotational and vibrational degrees of freedom on moving from the free components to the complexed system.[52,66,67] These calculations confirm our observations from the RMSD plots that thiazotropsin A stiffens DNA on binding, and therefore invokes the entropic penalty that the ITC studies indicate. The enthalpic term combines the molecular mechanics energy (E_{MM}) with the solvation energy (ΔG_{sol}) to give -49.53 kcal mol^{-1} and indicates the net exothermic drive for binding that our experimental studies demonstrated. At the molecular level, the stiffening of the macromolecule is reflected by a gain in steric or strain energy for the bonded component BADH, and confirms the total adaptation energy penalties for both ligand and DNA delineated in Table 6. Both the negative van der Waals and electrostatic terms suggest many favorable non-bonded interactions are formed between thiazotropsin A

Table 6 Parsing of the binding and adaptation free energies into the individual enthalpic, solvation and entropic components for the binding of thiazotropsin A with DNA using the separate trajectory approach with a polarizable force field. The binding energy components are the average values for both conformation populations (All_conf.) represented in Table 5. All values are in kcal mol^{-1}

Energy component	Total binding energy	Total adaptation energy of DNA	Total adaptation energy of ligand dimer
BADH	17.4	25.1	−8.4
VDW	−101.5	8.4	7.9
ELEC	−1145.7	6.7	12.8
PB	1189.3	−3.9	−1.8
SA	−9.1	0.2	0.2
ΔE_{MM}	−1229.8	40.2	12.4
ΔG_{sol}	1180.2	−3.7	−1.6
$-T\Delta S$	34.2	5.5	10.0

and the DNA dodecamer, a feature characteristic of minor groove binding ligands.[19] Again, these values must be viewed in association with the total adaptation energies (Table 4); the non-bonded penalties associated with conformational rearrangements undertaken during the binding event are more than compensated for by the net formation of new interactions between the ligand and the macromolecule. We suggest that the magnitude of the van der Waals contributions arises in part through our premeditated incorporation of larger alkyl groups in the ligand to promote lipophilic interactions with the groove walls.[20–23] The favorable electrostatics arise through a combination of the ten hydrogen bonds formed between the ligand dimer and the base pair edges on the groove floor (Table 3) and the two cationic DMAP tails interacting with the negative electrostatic potential of the DNA-phosphate backbone, which when combined with solvation show a clear net enthalpic drive towards complexation. Studies are currently underway with related analogues of thiazotropsin A binding to different oligodeoxynucleotide sequences to gain a fuller understanding of the key structural elements in both ligand and nucleic acid that drive association. We are performing these simulations using both explicit and implicit solvent representations in order to establish how to raise the throughput of the modelling without compromising accuracy (manuscript in preparation). Furthermore, extended simulation times will measure the impact of the parmbsc force field parameters on the free energies we generate.[68]

Conclusion

When developing new MGBs that bind selectively to DNA, it is essential to understand the factors that work both in favor of and counter to ligand binding. We have prepared a large library of MGBs that seek to recognize both the hydrogen bonding capacity of the groove floor to achieve specificity, and to exploit the lipophilic nature of the groove walls to enhance affinity. By way of this process, we found that the introduction of a C-alkyl into the thiazole ring of our ligand thiazotropsin A offset side-by-side binding and extended the reading frame of ligand from four to six base pairs. In 2004, we reported the first, well-characterized example of this effect in our detailed NMR study between the DNA duplex $d(CGACTAGTCG)_2$ with thiazotropsin A. To relate these structural examinations to binding energies, we have investigated the thermodynamics of thiazotropsin A interacting with the same sequence experimentally by ITC. Although thiazotropsin A has a greater enthalpic drive, it suffers from an unfavorable entropy that is uncharacteristic of other MGBs and is more analogous with intercalative binding. The entropic cost associated with intercalation is most likely through the rigidification of the DNA helix, which is less evident for MGB binding, particularly in the 1 : 1 complexes. However, our MD simulations suggest that the 2 : 1 complex formed by thiazotropsin A is significantly stabilized, which is reflected by reduced RMSD fluctuations compared with the free DNA dodecamer and through the loss in translational, rotational and vibrational degrees of freedom on moving from the free components to the complexed system that our normal mode analysis revealed. With respect to these MD simulations, we have demonstrated that both the AMBER polarizable (ff02) and non-polarizable (ff03) force fields can reproduce the structural characteristics of thiazotropsin A binding with the target nucleic acid we previously solved by NMR. In terms of absolute binding free energies, the determination of adaption free energies clearly demonstrates that a significant energetic penalty is associated with binding that is a consequence of the conformational rearrangement of the DNA dodecamer to accommodate side-by-side binding by the ligand. As a result, using the MM-PBSA methodology in a single trajectory approach that assumes such conformational changes do not take place is not compatible with this system, which is reflected by the significant overestimations of binding free energies using both force fields. The separate trajectory approach, particularly for the polarizable force field, returned binding free energies that approximate well with our experimental data. Our simulations also reveal that non-bonded interactions feature strongly in the association that work to overcome the unfavorable adaptation energies and entropy that are a consequence of complex formation. Perhaps the incorporation of structural features that seek to maximize lipophilic interactions with the groove walls to enhance the enthalpic drive will always be counter-balanced by unfavorable entropies that arise from the helical rigidification of the DNA.

Our previous work has focused on designing drug-like molecules with high affinity for DNA while maintaining sufficient sequence selectivity to have useful therapeutic effects. With this objective in mind, we have increased the hydrophobicity of our ligands through the incorporation of alkyl groups larger than methyl to balance the hydrophilicity that the hydrogen bonding amides and cationic tails impart on the molecules. Our study demonstrates that such modifications can indeed improve the binding enthalpy. The challenge we

now face is to see whether the opposing entropic contribution is an inherent feature of these modifications, or whether we can use this knowledge to our advantage, and find structural adjustments that harness entropy in our favor. Further studies are underway towards this end.

Acknowledgements

We thank the Royal Golden Jubilee PhD Program (3C.KU/47/B.1) and the Thailand Research Fund for support of WT, EPSRC and the Scottish Funding Council for their funding of the Physical Organic Chemistry initiative at Strathclyde, and both Michael Feig, of Michigan State University and the National Grid Service for access to cluster facilities.

References

1 R. Burnett, C. Melander, J. W. Puckett, L. S. Son, R. D. Wells, P. B. Dervan and J. M. Gottesfeld, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11497–11502.
2 P. B. Dervan, A. T. Poulin-Kerstien, E. J. Fechter and B. S. Edelson, *DNA Binders and Related Subjects*, 2005, Springer, Berlin–Heidelberg, vol. 253, pp. 1–31.
3 N. G. Nickols and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10418–10423.
4 P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284–289.
5 J. W. Lown, *Drug Dev. Res.*, 1995, **34**, 145–183.
6 S. Neidle, *Nat. Prod. Rep.*, 2001, **18**, 291–309.
7 U. Pindur, M. Jansen and T. Lemster, *Curr. Med. Chem.*, 2005, **12**, 2805–2847.
8 M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 1376–1380.
9 M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 7586–7590.
10 D. J. Patel, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 6424–6428.
11 W. S. Wade, M. Mrksich and P. B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 8783–8794.
12 J. G. Pelton and D. E. Wemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 5723–5727.
13 C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan and D. C. Rees, *Science*, 1998, **282**, 111–115.
14 S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird and P. B. Dervan, *Nature*, 1998, **391**, 468–471.
15 S. White, J. M. Turner, J. W. Szewczyk, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1999, **121**, 260–261.
16 A. N. Lane and T. C. Jenkins, *Q. Rev. Biophys.*, 2000, **33**, 255–306.
17 V. K. Misra and B. Honig, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 4691–4695.
18 I. Haq, J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins and J. B. Chaires, *J. Mol. Biol.*, 1997, **271**, 244–257.
19 J. B. Chaires, *Arch. Biochem. Biophys.*, 2006, **453**, 26–31.
20 N. G. Anthony, K. R. Fox, B. F. Johnston, A. I. Khalaf, S. P. Mackay, I. S. McGroarty, J. A. Parkinson, G. G. Skellern, C. J. Suckling and R. D. Waigh, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 1353–1356.
21 N. G. Anthony, B. F. Johnston, A. I. Khalaf, S. P. MacKay, J. A. Parkinson, C. J. Suckling and R. D. Waigh, *J. Am. Chem. Soc.*, 2004, **126**, 11338–11349.
22 A. I. Khalaf, A. H. Ebrahimabadi, A. J. Drummond, N. G. Anthony, S. P. Mackay, C. J. Suckling and R. D. Waigh, *Org. Biomol. Chem.*, 2004, **2**, 3119–3127.
23 N. G. Anthony, D. Breen, J. Clarke, G. Donoghue, A. J. Drummond, E. M. Ellis, C. G. Gemmell, J. J. Helesbeux, I. S. Hunter, A. I. Khalaf, S. P. Mackay, J. A. Parkinson, C. J. Suckling and R. D. Waigh, *J. Med. Chem.*, 2007, **50**, 6116–6125.
24 N. Spackova, T. E. Cheatham, F. Ryjacek, F. Lankas, L. van Meervelt, P. Hobza and J. Sponer, *J. Am. Chem. Soc.*, 2003, **125**, 1759–1769.
25 S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco and C. A. Laughton, *J. Am. Chem. Soc.*, 2001, **123**, 12658–12663.
26 H. Wang and C. A. Laughton, *Methods*, 2007, **42**, 196–203.
27 J. Dolenc, C. Oostenbrink, J. Koller and W. F. van Gunsteren, *Nucleic Acids Res.*, 2005, **33**, 725–733.
28 S. B. Singh, Ajay, D. E. Wemmer and P. A. Kollman, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 7673–7677.
29 H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin and V. S. Pande, *J. Chem. Phys.*, 2005, **123**, 084108.
30 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
31 J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401–9409.
32 V. Tsui and D. A. Case, *J. Phys. Chem. B*, 2001, **105**, 11314–11325.
33 C. M. Reyes and P. A. Kollman, *J. Mol. Biol.*, 2000, **297**, 1145–1158.
34 M. Lepsik, Z. Kriz and Z. Havlas, *Proteins: Struct., Funct., Bioinform.*, 2004, **57**, 279–293.
35 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
36 J. M. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.
37 A. Weis, K. Katebzadeh, P. Soderhjelm, I. Nilsson and U. Ryde, *J. Med. Chem.*, 2006, **49**, 6596–6606.
38 E. Vladimirov, A. Ivanova and N. Rösch, *J. Phys. Chem. B*, 2009, **113**, 4425–4434.
39 V. Babin, J. Baucom, T. A. Darden and C. Sagui, *Int. J. Quantum Chem.*, 2006, **106**, 3260–3269.
40 L. J. Yang, C. H. Tan, M. J. Hsieh, J. M. Wang, Y. Duan, P. Cieplak, J. Caldwell, P. A. Kollman and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 13166–13176.
41 P. Cieplak, J. Caldwell and P. Kollman, *J. Comput. Chem.*, 2001, **22**, 1048–1057.
42 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.
43 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
44 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
45 J.-P. Ryckaert, G. Ciccoti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
46 R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.*, 1989, **6**, 655–667.
47 K. Chin, K. A. Sharp, B. Honig and A. M. Pyle, *Nat. Struct. Biol.*, 1999, **6**, 1055–1061.
48 V. Babin, J. Baucom, T. A. Darden and C. Sagui, *J. Phys. Chem. B*, 2006, **110**, 11571–11581.
49 C. Laughton and B. Luisi, *J. Mol. Biol.*, 1999, **288**, 953–963.
50 I. Y. Torshin, I. T. Weber and R. W. Harrison, *Protein Eng.*, 2002, **15**, 359–363.
51 H. Naghibi, A. Tamura and J. M. Sturtevant, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 5597–5599.
52 J. B. Chaires, *Biophys. Chem.*, 1997, **64**, 15–23.
53 A. A. Gorfe and I. Jelesarov, *Biochemistry*, 2003, **42**, 11568–11576.
54 S. A. Shaikh, S. R. Ahmed and B. Jayaram, *Arch. Biochem. Biophys.*, 2004, **429**, 81–99.
55 I. Hamdan, G. G. Skellern and R. D. Waigh, *Nucleic Acids Res.*, 1998, **26**, 3053–3058.
56 D. R. Bevan, L. P. Li, L. G. Pedersen and T. A. Darden, *Biophys. J.*, 2000, **78**, 668–682.
57 T. E. Cheatham III and M. A. Young, *Biopolymers*, 2001, **56**, 232–256.
58 T. E. Cheatham and P. A. Kollman, *J. Mol. Biol.*, 1996, **259**, 434–444.
59 K. Boehncke, M. Nonella and K. Schulten, *Biochemistry*, 1991, **30**, 5465–5475.
60 J. Lah and G. Vesnaver, *J. Mol. Biol.*, 2004, **342**, 73–89.
61 H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891–913.

62 W. Wang, W. A. Lim, A. Jakalian, J. Wang, J. M. Wang, R. Luo, C. T. Bayly and P. A. Kollman, *J. Am. Chem. Soc.*, 2001, **123**, 3986–3994.
63 S. H. Huo, J. M. Wang, P. Cieplak, P. A. Kollman and I. D. Kuntz, *J. Med. Chem.*, 2002, **45**, 1412–1419.
64 H. Gohlke and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 238–250.
65 T. Lazaridis, *Curr. Org. Chem.*, 2002, **6**, 1319–1332.
66 R. S. Spolar and M. T. Record, *Science*, 1994, **263**, 777–784.
67 I. Haq, T. C. Jenkins, B. Z. Chowdhry, J. S. Ren and J. B. Chaires, *Methods Enzymol.*, 2000, **323**, 373–405.
68 A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.